DOCUMENT RESUME

ED 410 266 TM 027 048

AUTHOR Roeber, Edward D.
TITLE The Technical and Practical Challenges in Developing Innovative Assessment Approaches for Use in Statewide Assessment Programs.
PUB DATE Mar 97
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Course Content; Curriculum Development; Educational Assessment; Educational Change; *Educational Innovation; Elementary Secondary Education; Generalization; Performance Based Assessment; Sampling; Scaling; Staff Development; Standards; *State Programs; *Test Construction; Testing Problems; *Testing Programs
IDENTIFIERS *Council of Chief State School Officers; *Large Scale Assessment

ABSTRACT

Some of the reasons efforts are being made to reform education are discussed, and how these reforms are likely to affect U.S. schools is explored. Data collected annually by the Council of Chief State School Officers (CCSSO) indicate that almost all states have some form of assessment that is administered to all students at one or more grade levels across the state. In addition, in 1995-96, almost all states had developed, or were in the process of developing, content standards defining what students should know and be able to do. Widespread belief that schools are not helping all students achieve at the levels they are capable of reaching has spurred reform efforts. Student assessment is at the top of the list of things to reform, since it is considered a way to set more appropriate targets for students and to focus staff development and curriculum reform. New content standards may require new assessment methods, whether short-answer, open-ended, extended-response, or other innovative forms, including performance based assessment. Some of the most severe challenges states face in implementing innovative assessment are due to the practical aspects of large-scale, statewide testing programs. Technical challenges such as scaling, reporting, generalizability, and sampling issues must also be considered. The CCSSO has undertaken activities to develop new types of assessments and to engage in research about performance assessment. Particularly promising are approaches that coordinate assessment at the state, district, and classroom levels. The CCSSO is a leader in studying this type of coordination. (Contains 14 references.) (SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

# The Technical and Practical Challenges in Developing Innovative Assessment Approaches for Use in Statewide Assessment Programs

Edward D.Roeber
Council of Chief State School Officers

ED 410 266

## Introduction

Currently, much discussion is taking place about the quality of American schools, the skills needed by students, and the ways we should be assessing these achievements. Student assessment is viewed nationally as the pivotal piece around which school reform and improvement in the nation's schools turns. For example, student assessment is the key piece of Goals 2000, as well as other federal legislation such as the Improving America's Schools Act (IASA).

The result is that substantially more assessment is likely to occur in our nation's schools, and to take place in areas traditionally not assessed (such as the arts), using assessment strategies (such as performance assessments and portfolios) not typically used. States and local districts are reconsidering the models for systems of assessment and how assessment at the state and local levels can be coordinated to achieve the reforms desired in education. This digest lays out some of the reasons for the reform of assessment and how these reforms will likely affect our schools.

From data collected annually by the Council of Chief State School Officers (CCSSO) and the North Central Regional Educational Laboratory (NCREL), almost all states have some form of assessment that is administered to all students at one or more garde levels across the state (CCSSO/NCREL, 1996). In data reflecting the programs operated in 1995-96, virtually every state had developed or were developing content standards defining the knowledge and skills that students should know and be able to do. Many states were also in the process of revising their assessment programs to reflect these standards.

## Why is School Reform Occurring?

Widespread belief that schools are not helping all students achieve at the levels that they are capable of, nor that is needed, has spurred efforts to reform our schools. Concerns have been raised that the ways we teach students, as well as assess them, do not lead students to acquire needed knowledge or skills, nor help them apply and use their knowledge and skills appropriately. At the national and state levels, content standards containing the types of knowledge, skills, and behaviors now believed needed for all students to achieve at high levels are being developed. Starting with such efforts as the National Council of Teachers of Mathematics' Curriculum and Evaluation Standards for School Mathematics (NCTM, 1989), content standards are being developed in the arts, civics, economics, English, foreign languages, geography, health education, history, physical education, science, and social studies.

TM027048

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Edward Roeber
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

School reform is also motivated by the belief that there are competencies needed for graduates to enter the workforce successfully. The Secretary's Commission on Achieving Necessary Skills developed a list of generic competencies and foundation skills that all workers will need in the future (U. S. Department of Labor, 1991). Such skills include flexible problem solving, respecting the desires of the customer, working well on teams, taking responsibility for one's own performance, and continuous learning. These skills have been developed to guide the efforts of educational reform in the direction helping more students to make the transition to work successfully.

Collectively, these standards represent substantial challenges for the American schools. They imply that all students will need to achieve at much higher levels. New strategies for assessment are also implied by these content standards.

**How Reform of Assessment Fits School Reform?**

Student assessment is at the top of the list of things to tinker with by policymakers at the national and state levels, since it is viewed as a means to set more appropriate targets for students, focus staff development efforts for the nation's teachers, encourage curriculum reform and improve instruction and instructional materials in a variety of subject matters and disciplines (Darling-Hammond & Wise, 1985). Assessment is an important part of the equation because it is widely believed that what gets assessed is what gets taught, and that the format of assessment influences the format of learning and teaching (O'Day & Smith, 1993). The hope of policymakers is that the changes in assessment will not only bring about the needed changes in students, but also in ways schools are organized (Linn, 1987; Madaus, 1985). Interest in performance assessment has also been justified on the basis that using such measures will accomplish (or at least promote) educational equity (National Center on Education and the Economy, 1989). Student assessment carries a heavy load these days!

Of course, outside pressure from external testing programs can be ignored or resisted by local educators (Smith and Cohen, 1991). There is also ample evidence of the distortions in teaching that external testing programs can create (Shepard & Smith, 1988). Rather than encourage reform of teaching, inappropriate teaching to the test may occur (as opposed to teaching to the broader domain covered by the test). Rather than creating opportunities for all students to learn to high levels, even new forms of assessment may lead to tracking and limiting opportunities for some students (Darling-Hammond, 1994; Oakes, 1985).

Assessment reform should occur along with professional development, instructional development, and other strategies designed to assure that all of the changes are mutually supported. Coordination of assessment reform at the national and state levels with assessments at the local level is also important, so that each will present a coherent view of student performance, not simply be "stuck" together.

**Types of Assessments**

An essential element of the design of assessment is the choice of exercise

type(s). New content standards may require different assessment methods. Among the assessment techniques now being considered are short-answer, open-ended; extended-response, open-ended; individual interview; individually- or group-administered performance events; individual or group performance tasks in which students have extended time; projects; portfolios; observations of students; and anecdotal records, in addition to multiple-choice exercises. A broader repertoire of techniques is increasingly being used.

**Useful Assessment Designs**

Typically, student achievement is measured with available student test data, often using information from district or state testing programs. Information collected less formally at the classroom levels is not typically included in school improvement plans, even though such information could provide valuable insights into student learning.

The nature of information needs should form the basis for an assessment design. In a top-down model, policymakers develop an assessment design that meets their needs, hoping the data may be useful by persons at lower levels. An alternative is to build the assessment system needed at the local level, aggregating the information upwards to the district, state and national levels.

Another model, based on the assumption that multiple approaches will allow different users' needs to be met, is to develop comprehensive assessment system using different assessment formats to meet different users' needs. Various assessment strategies can be implemented together at the different levels to provide for the different information needs in a coordinated, coherent manner (Darling-Hammond, 1994).

For example, local districts can adopt a portfolio system for improving instruction, while the state carries out matrix-sampling across important standards. The information collected by the state can become part of the student's portfolio, thereby strengthening the quality of the information contained in students' portfolios. The state could also provide opportunities for teachers to learn to score the open-ended written and performance assessments, thereby enhancing teachers' capabilities of observing and rating student performances in their classrooms.

In this case, the elements of the assessment system at the different levels build on and support the elements at the other levels. It is also anticipated that information collected at the different levels can be reported in a more understandable manner, since the same standards apply in different ways. This assessment model enhances the reforms of schools so many desire.

**Practical Challenges Inherent in Using New Forms of Assessment**

Some of severest challenges that states face in implementing innovative approaches to assessment are due to the practical aspects of statewide testing programs. For example, programs that desire to use more constructed-response assessment exercises find that such types of exercises

are more time-consuming, so that either they require more testing time overall (perhaps more than schools will permit to be devoted to such external testing programs) or the technical quality of the data may be compromised by using too few items for the types of reports of assessment results to be provided. This is true particularly for programs in which detailed individual student results or sub-scores of the various standards that comprise the assessment will be reported. However, the value of the external assessment program may be judged on the utility of the data that is returned to parents and teachers; the more specific the information (by student and/or sub-skill), the better in the view of teachers and parents.

The testing time limits (which do vary from one state to another) can impact the quality of the assessment in another way: the breadth of the assessment. Since innovative assessment exercises take more time per exercise for students to complete, if testing time is not expanded, the number of exercises used may be cut sharply with the result that fewer aspects of the state's content standards are assessed (even if the assessments are more authentic). While "teaching to the test" when the assessment is more authentic may result in more authentic "practice," the impact of reduced "coverage" of the standards could be a narrower curricular focus to the assessment (and instruction prompted by the assessment) than might have been encouraged by an assessment that was primarily comprised of multiple-choice items.

In a similar fashion, the time that it takes to return results to schools (and whether the data is returned in a timely manner so as to be useful in either remediation or instructional improvement) is also a practical impediment of innovative approaches to assessment. Constructed response exercises take time to score. Tests comprised of multiple-choice items can be scored and reported, even for large-scale programs, in a matter of days. However, it is more typical for large-scale constructed response assessments to take up to several months to return results for individual students. This time lag between assessment and reporting is so large that local educators may view the results (and the overall assessment program) as relatively useless, since the results come back to them so far after testing that the results can not be relied on; the results of spring testing may not be returned until the following school year, when some of the students have moved, and the remaining students are dispersed to various schools and classrooms around the district. One way to reduce this turn-around time is to have classroom teachers score the tests, but this is not a popular activity for local educators.

**Technical Challenges Inherent in Using New Forms of Assessment**

There are also a number of technical challenges inherent in new forms of assessment, some imposed by the practical constraints covered above, and others inherent in these forms of assessment. These need to be addressed as well. Some of these challenges are as follows:

- Scaling Issues There are a couple aspects of the scaling issue that must be attended to when performance assessments are used. Due to practical constraints, the numbers of constructed response items used to assess a particular subject area may be few in number, and they may be purposely selected to scatter across a set of standards, with only one exercise use to represent any area of the assessment. Hence, the items

may be intentionally selected to replicate one another, which is then a challenge in scaling the items.

When a "mixed" assessment model is used, an additional challenge of scaling multiple-choice and constructed response items together can be introduced.

Finally, there can be the issue of whether special needs students' (English-language learners and students with disabilities) performances can be placed on the same scale as other students. Did the accommodations change the nature of the assessments given to these students? Were special needs students who did not receive accommodations assessed in a appropriate manner?

All of these are scaling issues that must be considered when determining the manner in which items can be aggregated.

- Reporting Issues Just as there are additional scaling issues in the use of performance assessments, there are also reporting issues of a similar nature. This really represents an inherent conflict: performance assessment exercises are used because they may well contribute a unique understanding of students, yet the goal is to report them together with the multiple-choice items used on one overall report of student performance. If the performance assessments fail to contribute substantial unique variance in overall student performance, are they worth the considerable investment of classroom time and money to use them? Yet, if they are too unique, can they be reported on the same scale as the multiple-choice items? An interesting conundrum!

- Generalizability Since few performance assessment exercises are used in a typical assessment program, can the few exercises selected truly represent curricular domains that sometimes are quite broad? Past research evidence indicates that it is difficult for the performance on one or a few exercises to generalize to other, supposedly comparable assessment exercises. How stable are the estimates of student performance?

- Reporting Trend Data Another major reporting need is to report student performance over time. As the assessment program is used over time, there is a natural desire at both the state and local levels to examine whether achievement is improving. This means that assessments in which performance assessments are used and where the generalizability of these assessments is lower than for the multiple-choice sections of the assessments, the challenge of longitudinal reporting will need to be met. Is the form of assessment used each year sufficiently equivalent that observed differences between student performance are not due to the instruments used?

- Use of Matrix Sampling One way to broaden total coverage of the assessment without increases in testing time per student is to administer only a sample of the assessments to a student by breaking the assessment into several pieces and giving each student only one piece of it. However,

when the assessment must produce individual student results in some manner, this introduces the challenge of assuring that the results are comparable across students (that is, that it does not matter which form of the assessment is assigned to any student). This means that some form of scaling and/or equating (adjacent form or using a common set of assessment exercises across all forms) will need to be used. This adds additional challenges to the scaling and reporting challenges detailed above.

- Standard-Setting The use of performance assessments, which often are based on multifaceted frameworks or content standards, may introduce the need for different methods for setting standards. The more traditional methods (modified Angoff or the contrasting groups approaches) may not work as well when the assessment is based on several traits and uses two or more assessment methods.

- Consequences One major reason for using performance assessments is that they will change the manner in which students are taught, as well as the amount of learning shown by students. While these a fairly hefty requirements to lay on any form of assessment, they are oft-ascribed reasons for including performance assessments as at least a portion of the overall assessment program. However, what evidence can be gathered that such changes and improvements are actually occurring? Are the changes observed worth the costs of teaching time and monies? Can such changes only occur by using performance assessments?

**CCSSO Research and Development Activities**

The Council has undertaken activities both to develop performance assessments and other innovative approaches to assessment in several areas, as well as to engage in research about performance assessment. This work is being done by the State Collaborative on Assessment and Student Standards (SCASS) projects in several areas. Developmental work in the SCASS science, health education, and social studies projects includes the development of modules of multiple-choice and constructed-response exercises, hands-on performance events, and instructionally-imbedded activities that students may work on individually or in small groups for several weeks outside of class. The SCASS arts and workplace readiness group is developing prototype performance assessments. In the SCASS science, health education, and primary-level assessment projects, portfolio assessment strategies and materials have also been developed. These projects show that it is practical and cost-effective for state assessment programs to develop performance assessments, some of which states will use and others will be used by local school districts within the member states.

In addition to these developmental activities, the Council is engaged in a federally-funded research project called the SCASS Technical Guidelines for Performance Assessment project[1]. Twenty-two states belong to this group, which is funded by an OERI Assessment Development and Evaluation grant. Members work with expert in relevant areas to design and implement

[1] Some of this text is taken from the "Request for Proposals" which the project uses to engage researchers in conducting studies for the group. The author of the RFP is Phoebe Winter, CCSSO.

research that will foster the development of technically-sound performance assessments. The TGPA project is a unique collaboration of educational researchers and state assessment staff, resulting in practical, well-executed research useful both in advancing the field as a whole and in meeting the immediate needs of large-scale assessment programs. Top researchers use the states' databases on which to conduct their research, which permits replications of research across different types of programs used for different purposes. It also helps to assure the applicability of research findings to the several types of assessments typically used by different states.

Currently, this SCASS project is sponsoring development or implementation research on the following topics:

- Validation of Performance-Based Assessments
- Linking Scores on Alternative Assessments
- Setting Performance Standards in Alternative Assessment
- Issues in Combining and Reporting Scores from Various Types of Assessments
- Empirical Evidence for Informing Assessment Accommodation Decisions
- Assessment of Limited-English Proficient Students
- Producing Reliable Student-Level Scores Using Matrix-Sampling
- Obtaining Multiple Scores from a Single Assessment
- Using Performance-Based Assessment Data to Measure School Progress

In SCASS TGPA-sponsored research, each study is coordinated by a study group consisting of member states participating in the overall SCASS project, plus others interested in the area of research being studied. Study group members work with researchers to review and approve the research design and project budget, plan the conduct of the study, and develop a plan for disseminating the results of the research. The research design is presented in the form of a detailed proposal, which includes a description of the research methodology, a review of recent relevant literature, the timeline for the research, and the responsibilities for participating states and local school districts for data collection, data analysis, and reporting.

Once the research proposal is approved, the researcher(s) carry out the study, working with the study group chair and committee. At appropriate points, the researcher(s) reports to the TGPA group at one or more of their quarterly meetings. Once completed, the researcher(s) report the results to the full group. The results are also used y the TGPA group to develop a "practical guidelines" book. The group is writing this book in order to indicate the ways in which performance assessments can be feasibly used in state and local assessments. It is the ultimate goal of the project to encourage the use of performance assessments on a wider scale by addressing the challenges that serve as impediments to the use.

## Summary

This is indeed a time when American schools are being challenged to provide opportunities for students to achieve at much higher levels. Assessment is viewed as one of the essential elements in assisting schools to address the standards now deemed to be important in a manner that will help all students to achieve them. The major challenge for assessment is to implement these additional assessments in a coordinated manner so that the amount of assessment is supportive of the changes needed, not overly burdensome to teachers or students. Models for coordinating assessment at the state, district and classroom levels appear most promising. Yet, such models introduce considerable practical and technical challenges into assessment program designs. Fortunately, there are feasible ways of addressing these issues, and CCSSO is a leader among organizations addressing these issues and helping states and local school districts improve the quality of their assessments.

**References**

Council of Chief State School Officer/North Central Regional Educational Laboratory, "Annual Report of State Student Assessment Programs." Washington, DC and Chicago, IL: Author. 1996.

Darling-Hammond, L. and A. Wise (1985). "Beyond Standardization: State Standards and School Improvement." Elementary School Journal, 85, 315-336.

Darling-Hammond, L. (Spring, 1994). "Performance Assessment and Educational Equity." Harvard Educational Review, 64 (1): 5-29.

Linn, R. (1987). "Accountability: The Comparison of Educational Systems and the Quality of Test Results." Educational Policy, 1(2): 181-198.

Madaus, G. (1985). "Public Policy and the Testing Profession - You've Never Had it So Good?" Educational Measurement: Issues and Practices, 4 (1): 5-11.

National Center on Education and the Economy (1989). To Secure Our Future: The Federal Role in Education. Rochester, NY: Author.

National Council of Teachers of Mathematics. Curriculum and Evaluation Standards for School Mathematics. Reston, VA: Author. 1989.

Oakes, J. (1985). Keeping Track: How Schools Structure Inequality. New Haven: Yale University Press.

O'Day, J.A. and M. Smith (1993). "Systemic School Reform and Educational Opportunity." In S. Fuhrman (Ed.), Designing Coherent Educational Policy: Improving the System. San Francisco: Jossey-Bass, pages 250-311.

Roeber, E. (1992). "1. Developing the Comprehensive Assessment System: A. Top Down, B. Bottom Up, C. Both, D. None of the Above." Paper Presented at the 1992 Education Commission of the States Conference on Large-Scale Assessment.

Rudner, L. and C. Boston (Winter, 1994). "Performance Assessment." The ERIC Review, 3 (1): 2-12.

Shepard, L.A. and M.L. Smith (1988). "Escalating Academic demand in kindergarten: Counterproductive Policies." Elementary School Journal, 89: 135-145.

Smith, M. and M. Cohen (September, 1991). "A National Curriculum in the United States?" Educational Leadership, 49 (1): 74-81.

U.S. Department of Labor. Secretary's Commission on Achieving Necessary Skills. Washington, D.C.: Government Printing Office. 1991.

The author of this paper is Edward D. Roeber, Director, Student Assessment Programs, Council of Chief State School Officers, One Massachusetts Avenue, N.W., Washington, D.C.